International Journal of Scientific & Engineering Research, Volume 4, Issue 2, February-2013
ISSN 2229-5518

Cloud Base Height Retrieval from Portable Automated Lidar Data by Wavelet Analysis

Gerry Bagtasa[1,2], Nobuo Takeuchi[3], Hiroaki Kuze[3]

Abstract— The Portable Automated Lidar (PAL) has been shown to be effective in observing cloud dynamics and aerosol optical properties in the troposphere. The continuous monitoring of the atmosphere, however, has indicated that the detection range is rather limited for the daytime data because of the background noise from the sky radiance. Here we apply the wavelet denoising method to improve the Signal-to-Noise Ratio (SNR) of the time-dependent lidar data. As a result, SNR increased by 7.9% when compared to data smoothing based on moving average. Moreover, the analysis of wavelet coefficients enables direct retrieval of the cloud base height. It is found through manual comparison that this method shows 4.4 times less false positive and 2.1 times less false negative detection on average, when compared to a conventional method such as the threshold method.

Index Terms— Lidar, Cloud Base Height, Portable Automated Lidar, Wavelet Analysis, Signal Processing, Clouds, Laser.

1 Introduction

Lidar is a useful tool in monitoring the atmosphere. Behavior of clouds and aerosols, in particular, can be elucidated by means of elastic, backscattering lidars. The portable automated lidar (PAL) developed by the Center for Environmental Remote Sensing (CEReS), Chiba University, has the capability of unattended, long-term operation. Lagrosas et al. [1] demonstrated the effectiveness of the PAL system by observing the oscillatory behavior of the aerosol layer height, vertical motion of the aerosol layer, and speed of rain drops. A good correlation was found between the intensity of backscattered light and suspended particulate matter concentration inside the boundary layer. Also, mass extinction efficiency (MEE), defined as the ratio of aerosol extinction to its mass concentration, was computed together with ground-based SPM measurements [2,3].

The lidar technique has good flexibility and holds promise for providing accurate measurements of wide variety of aerosols and meteorological parameters of the atmosphere [4]. However, noise inherent in the system during observation degrades the information embedded in the lidar signal. Due to the inverse square nature of the lidar data, signal to noise ratio of the system decreases with range. In practice, procedures are taken to improve the lidar signal. For instance, the received power $P(r, \lambda)$ at range r and wavelength λ is the average of 28000 (1.4 kHz in 20 s) laser pulse shots for a typical PAL signal. This process reduces most of the random interference and digitization noise of the system, but considerable noise will still affect the data especially at longer ranges. In addition, moving average is often employed to further improve lidar data. This process, however, cannot eliminate speck values, especially negative values produced by noise [5].

The purpose of this study is to use wavelet analysis to improve the quality of PAL data and use the wavelet transform to determine cloud occurrence and/or cloud base height (CBH). Since the mathematical complexity of wavelet analysis requires intensive and time consuming programming, the analysis in this study is based on an open source Matlab functions, called Wavelab, developed by Buckheit et al. [6].

2. Theory

2.1 Lidar Theory

The intensity of the backscattered beam is dependent on the backscattering properties of the atmosphere at wavelength λ and distance r, which is in turn, dependent on the quantity, size, shape and the refractive index of the particles causing the backscattering. The backscattered intensity also depends on the round trip atmospheric attenuation which characterizes the transmission properties of the atmosphere.

The range-corrected, backscattered signal $X(r, \lambda)$ is given by

$$X(r,\lambda) = P(r,\lambda) r^2 \tag{1}$$

where $P(r,\lambda)$ is the power scattered from a target. Although the backscattered signal inherently includes multiple scattering, a narrow FOV of the PAL systems (0.2 mrad) ensures the detection of mostly single scattering events. Since the PAL system takes slant path measurements with an elevation angle θ (38 deg), here we show the raw signal as $P(z,\lambda)$, where $z = r \sin\theta$ describes the height of the target.

2.2 Wavelet transform

The wavelet function $\psi_{s,\tau}(t)$ can be expressed in terms of the mother wavelet ψ by

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \tag{2}$$

where s and τ are scaling and translation parameters, respectively. The continuous wavelet transform is formally given by [7]

$$\gamma(s,\tau) = \int f(t) \psi_{s,\tau}(t)\, dt \tag{3}$$

where $f(t)$ is the function of the signal, which is decomposed to the basis function $\psi_{s,\tau}(t)$ called wavelets. The variables s and τ represent the new dimension after the wavelet transform. Conversely, the inverse wavelet transform is given by

$$f(t) = \int\int \gamma(s,\tau) \psi_{s,\tau}(t)\, d\tau ds \tag{4}$$

1. Institute of Environmental Science & Meteorology, University of the Philippines, Diliman, Quezon City, Philippines. E-mail: gerrybagtasa@gmail.com
2. Natural Sciences Research Institute, University of the Philippines, Diliman, Quezon City, Philippines.
3. Center for Environmental Remote Sensing, Chiba University, Chiba, Japan.

A wavelet function $\psi(t)$ is defined to satisfy the following conditions: (i) its square integral is finite, called the admissibility condition

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty \tag{5}$$

and (ii) its integral in the time domain is zero:

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0 \tag{6}$$

Continuous wavelet transform (CWT) cannot be practically computed using analytical equations. Therefore, we have the discrete wavelet transform (DWT). In CWT, there is no constraint in the value of s and τ, and they can in principle map the whole (s, τ) plane. In the DWT, however, the possible values of (s, τ) are restricted to

$$s = s_0^j \tag{7}$$

and

$$\tau = k\tau_0 s_0^j \tag{8}$$

where k and j are integers. The wavelet analysis of lidar signal in this study is based on orthonormal wavelet bases with $s_0 = 2$ and $\tau_0 = 1$. This is referred to as the dyadic dilations and translations.

2.3 Lidar signal denoising

Wavelet based denoising differs from traditional filtering approaches in that it is nonlinear due to a thresholding step. It employs thresholding in the wavelet domain and has been shown to be asymptotically near optimal for a wide class of signals corrupted by additive white Gaussian noise [8]. Denoising by thresholding involves the following steps: the first step is to perform a wavelet transform of a noisy data, second is to perform a thresholding where the threshold depends on the noise variance, and third, the coefficients obtained from step 2 are padded with zeros to produce a wavelet transform and this is inverted to obtain a denoised signal. There are two versions of thresholding, hard and soft threshold [9]. In hard thresholding, any coefficients less than or equal to the threshold are set to zero, whereas in soft thresholding, any coefficients less than or equal to the threshold are set to zero, then the threshold is subtracted from any coefficients greater than the threshold. In this study, hard thresholding is used so as to further discriminate cloud signals with the background noise. The advantage of wavelet-based denoising is that noise is largely suppressed while features in the original signal remain mostly unchanged, in contrast with traditional linear methods of smoothing which compromise noise suppression with the broadening of signal features.

Real lidar return signals tend to have a dominant low frequency component, and almost all of the high frequency component of the signal can be assumed to be noise [10]. Therefore, by reducing the value of the scale in the high frequency component, particularly $j = 8$ and $j = 9$ of the signal by hard thresholding method, noise in the signal can be suppressed in an effective way. Fang and Huang [10] demonstrated the advantage of wavelet denoising over a more traditional approach using Butterworth filter. This study follows the methodology of Fang and Huang, on the other hand, we use the simpler Daubechies wavelet (Sec. 4.1) and hard thresholding and extended the analysis to a new cloud retrieval method.

2.4 Cloud base height detection using wavelet coefficients

One aim of this study is to present a new method in cloud base height detection with the use of wavelet decomposition analysis. A cloud in a lidar profile is indicated by a sudden increase of backscattered lidar signal due to the high backscattering coefficient inherent to clouds. Traditionally, the threshold method is used to determine the presence of a cloud in a lidar profile [11]. Other methods include detection of zero crossing in the derivative of the signal [12] and compare lidar signals to archived clear sky profiles [13]. Following these methods, noise estimates and/or smoothing are used to reject false peaks from noise. In this study, threshold method is used to compare with the wavelet retrieval method. A threshold value is set to distinguish clouds from specks arising from noise, and any sudden increase in the signal intensity exceeding the threshold value is considered as a cloud. The threshold method in cloud determination is effective in cases where there is sufficient signal-to-noise ratio, but is ineffective in determining high altitude clouds where lidar signals tend to show lower SNR due to the background solar radiation [14]. A sudden increase in the backscattered signal is reflected in the wavelet coefficient in the time-scale domain. Larger wavelet coefficients in the time-scale domain corresponding to a cloud consistently appear in scale depths $j = 9$ to 5. Starting from the finest scale ($j = 9$), information about the different frequency components of the cloud is contained through $j = 5$, below which the resolution is too low to contain any details of a cloud. The detailed result is shown in Section 4.

3 PAL System

The portable automated lidar system is placed at Chiba Prefectural Environmental Research Center (CERC) (35.52N, 140.07E, about 40 km southeast of Tokyo). The PAL system is installed indoors at a height of about 4.5 m from the ground level. The laser beam is pointed toward the north sky with a fixed elevation angle of 38°. The distance between the lidar location and the seashore (east coast of the Tokyo bay) is about 2.5 km in the laser beam direction.

The main components of the PAL system are a pulse laser mounted on a 20 cm diameter telescope, a photo-multiplier detector, a signal scaler and a personal computer (PC) for data acquisition and system control. Table 1 summarizes the system specification. A detailed description can be found elsewhere [1]. In order to reduce the background due to the skylight during daytime, a narrow field-of-view angle of 0.2 mrad and a narrow filter bandwidth of 0.5 nm are employed. The laser beam is reflected by two prisms so that its axis coincides with the optical axis of the telescope. Automatic realignment is done every 15 min by adjusting the orientation of one of the prisms. The backscattered signal is collected by the telescope, detected by the photo-multiplier tube in photon-counting mode, averaged for 20 s, and stored in the PC.

The PAL system was continuously operated from December 2002 to December 2007, with occasional maintenance. For this study, a total of 30780 hours of data or approximately 5.5 million lidar profiles are used to determine the strength of wavelet decomposition analysis on lidar signals.

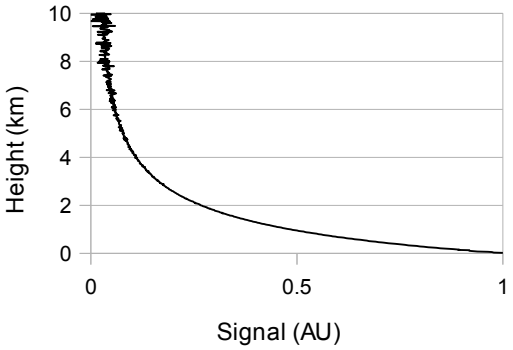
TABLE 1
SPECIFICATION OF THE PAL SYSTEM.

Transmitter		
Laser	LD-pumped Nd:YAG	Q-switch
Wavelength	532 nm	
Laser Pulse Width	50 ns	
Repetition Rate	1.4 kHz	
Laser Pulse Energy	15 µJ	
Laser Beam Divergence	50 µrad	
Receiver		
Telescope Diameter	20 cm	
Telescope Type	Cassegrain	
Field of View	0.2 mrad	
Filter Bandwidth	0.5 nm	
Detector	PMT (HPK-R1924P)	
Quantum Efficiency	10 - 25%	

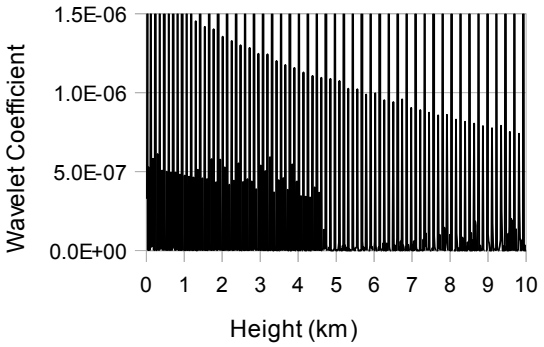
4. RESULTS AND DISCUSSION

4.1 Lidar data denoising

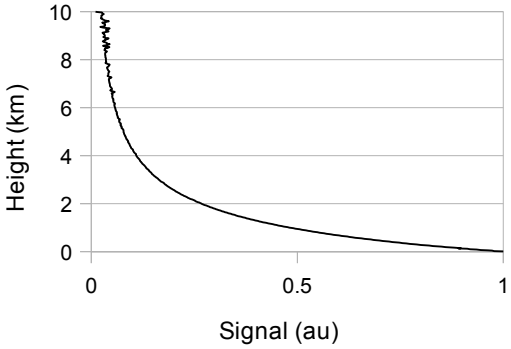
The Daubechies4 wavelet was selected for the wavelet decomposition. This wavelet provides a powerful tool for signal processing. The Daubechies wavelet is simple enough as it is defined in the same way as the Haar wavelet transform, which is the simplest wavelet transform, that is by computing the running averages and differences via scalar products with scaling signals and wavelets. Haar transform performs an average and difference on a pair of values, then shifts over by 2 values for the next pair. If a sudden change takes place from an even value to an odd value, the change will not be reflected in the high frequency coefficients. Daubechies, on the other hand, also shifts by 2 elements each step but averages and differences are calculated over 4 elements. This use of "overlapping windows" enables high frequency coefficient spectrum to reflect all high frequency changes, making this wavelet useful in signal compression and noise removal [15].



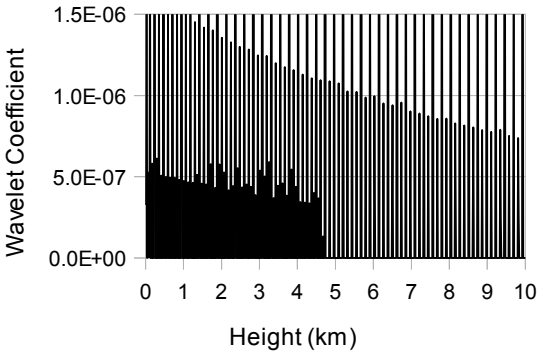
(a)



(b)



(c)



(d)

Figure 1. (a) shows a simulated lidar signal with additive white Gaussian noise and **(b)** is its corresponding time-scale domain. **(c)** is the wavelet denoised "noisy" data and **(d)** its corresponding time-scale domain graph.

Figure 1 (a) shows a simulated lidar profile (using extinction coefficient profile derived from LOWTRAN7 [16] and standard atmospheric model US76) with additive white Gaussian noise with zero mean and Fig. 1 (c) shows the same profile after removing the high frequency component. Figure 1 (b) and (d) show the corresponding dyadic sampling in the time-scale (frequency) domain for the noisy and denoised profiles, respectively. The horizontal axis, originally accounting for time, is converted to height for ease of the comparison with the lidar profile, while the vertical axis shows the wavelet coefficient of scale depth ranging from 0 to j-1, where higher frequency components lie closer to one another (j in the figures is equal to 10, corresponding to a dyadic signal length of 2^{10}=1024). Wavelets can be translated to $2^{(\text{scale depth})}$ points on each scale depth.

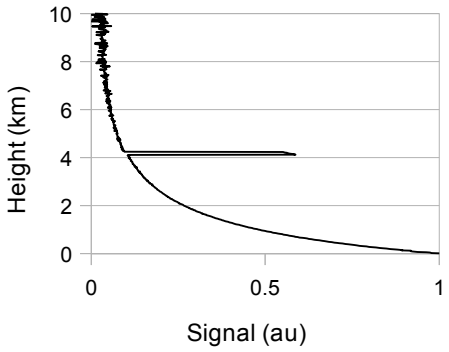
Theoretically, it is possible to choose a threshold value without *a priori* knowledge of the signal: a value at four times the standard deviation σ, for example, would eliminate 99.9% of the noise magnitude. Here, we set the threshold value at 4.5σ to further eliminate the noise signal energy. The effectiveness of the denoising is quantitatively measured by

$$SNR = 10\log\left(\frac{A_{signal}}{A_{noise}}\right)^2 \quad (9)$$

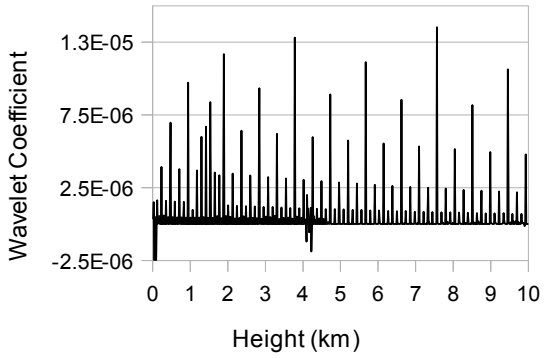
where A_{signal} and A_{noise} are the root mean square amplitude of the signal and noise, respectively. Comparing the SNR increase of real PAL data using wavelet denoising and moving average, the moving average produced an increase of 17.11 db SNR, average for all 5.5 million profiles, while wavelet based denoising produced an increase of 18.46 db, a 7.9% increase in SNR. Although this is a relatively modest increase, the preservation of features due to cloud peak intensities is an obvious advantage of the wavelet denoising, as discussed in the following.

4.2 Cloud Base Height determination

Figure 2 (a) shows a noisy lidar profile obtained by adding a thin cloud at approximately 4 km to the profile presented in the previous section. The corresponding diagram in the time-scale domain (Fig. 2 (b)) shows a distinct negative coefficient that indicates the sudden change in the signal profile due to the presence of the cloud. A real PAL data and it's time-scale domain graph shows similar characteristics as illustrated in Fig. 3 (a) and (b). Figure 4 (a) shows another simulated profile with cloud which is optically thinner and higher in altitude to simulate a low SNR condition and Fig. 4 (b) is the corresponding time-scale domain graph. Figure 5 shows a similar low SNR condition PAL data.

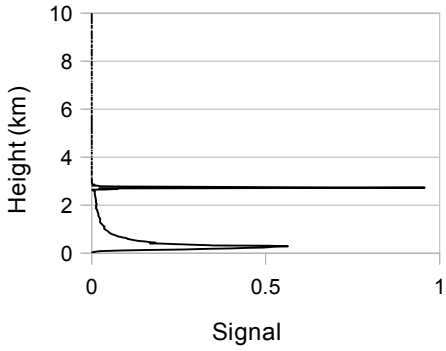


(a)

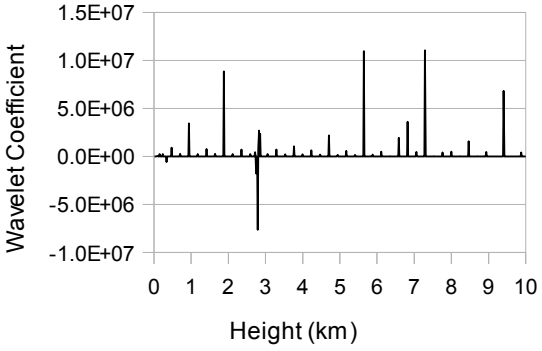


(b)

Figure 2. (a) A simulated lidar with high intensity backscattered signal simulating a cloud at 4 Km and **(b)** its corresponding time-scale graph.

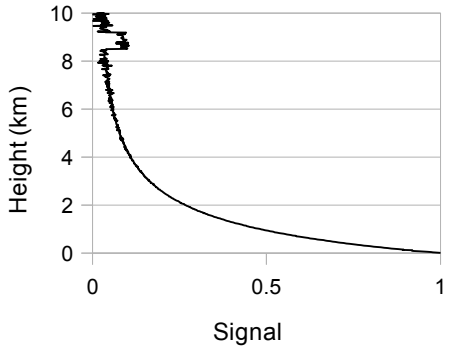


(a)

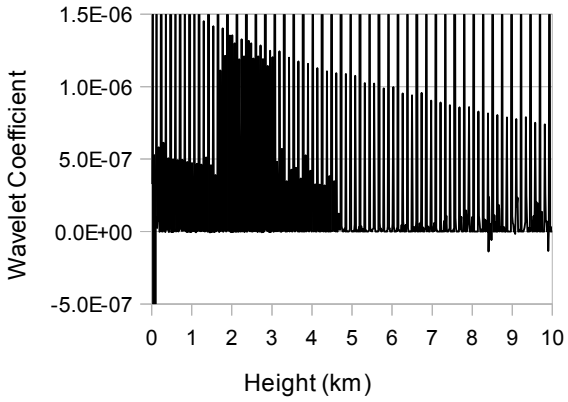


(b)

Figure 3. (a) A typical PAL profile data with cloud at approximately 2.9 km and **(b)** its corresponding time-scale graph

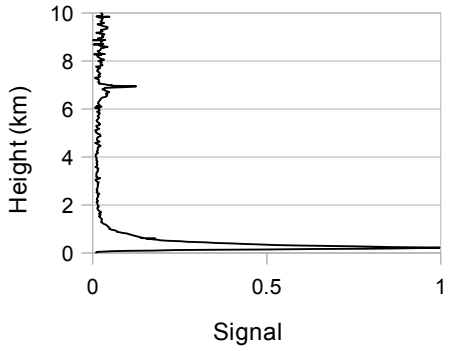


(a)

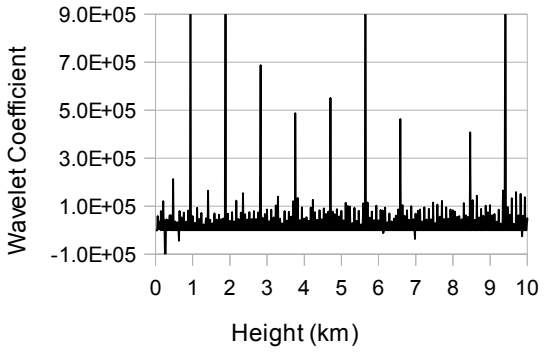


(b)

Figure 4. (a) A simulated lidar signal with optically thin cloud at 8.5 Km and **(b)** its corresponding time-scale graph.



(a)



(b)

Figure 5. (a) A typical PAL signal with optically thin cloud at 7 Km and **(b)** its corresponding time-scale graph.

Since there are no available CBH observations other than the PAL system, we validated the wavelet-based retrieval method and the threshold method with visual inspection of cloud occurrences. Here, we define false positive error as cloud detected by the algorithm without visually discernible clouds and false negative error as cloud not detected by the algorithm on visually apparent clouds. Relative to the total number of profiles with clouds derived from manual inspection, the wavelet analysis detected 3.5% false positives while the traditional threshold method yielded 15.2% false positive detections, giving 4.37 times more false positive detection for the entire 30780 hr PAL dataset. On the other hand, of all the derived cloud occurrences, there is 2.2% and 4.6% false negative detections by the wavelet and threshold method, respectively, indicating a difference of 2.1 times. Fig. 6 and 7 show a typical example of cloud base retrieval time height index (THI) graph by wavelet analysis and threshold method, respectively. Pal data spans from 1700H Japan Standard Time (JST) of 28 December 2004 to 0600H the next day. False positives are highlighted by the rectangle and false negatives are encircled in both figures (Fig. 6 and 7). Furthermore, false positive detections did not vary much for low (below 5 km) and high (above 5 km) altitudes. For the threshold method, cloud free conditions produced more false positive detections above 5 km while low altitude false positive detections appeared right above detected clouds, this is due to the severe attenuation of the laser beam by low lying optically thick clouds that degrades SNR above the clouds. Table 2 shows the summary of the comparison between the wavelet and threshold cloud retrieval methods.

TABLE 2. COMPARISON OF CLOUD DETECTION USING WAVELET ANALYSIS THE THRESHOLD METHOD.

	Wavelet	Threshold
False positive	3.48%	15.20%
False negative	2.20%	4.55%
Average False positive below 5 km	1.68%	7.09%
Average False positive above 5 km	1.80%	8.11%

Sample size: 30780 hours

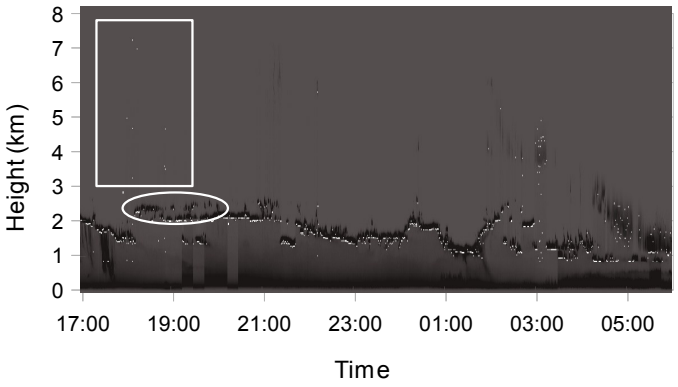


Figure 6 Cloud base retrieval by wavelet analysis THI PAL graph from 1700H JST of December 28 to 0600H of December 29 2004.

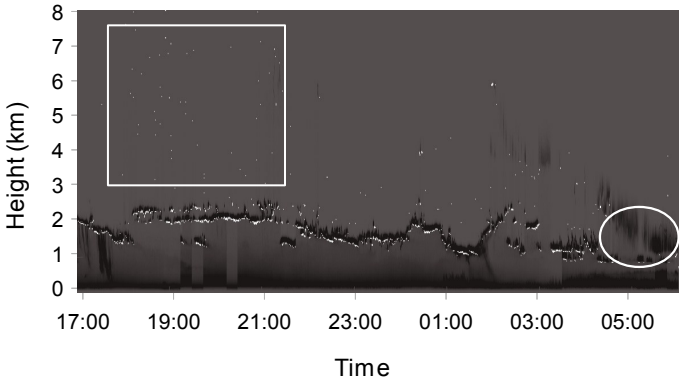


Figure 7. Cloud base retrieval by threshold method THI PAL graph from 1700H JST of December 28 to 0600H of December 29 2004.

5 Conclusion

We have described a cloud base height determination approach based on wavelet decomposition analysis. Wavelets have good time-frequency (scale) localization, which makes it better in representing non-repeating signals. At the same time, denoising using wavelet analysis provides a tool appropriate for cloud retrieval with less processing time. Only a single threshold value for denoising and cloud retrieval is needed for all profiles in the wavelet analysis method while threshold method ideally requires different threshold values depending on background noise conditions. The wavelet-based cloud retrieval algorithm presented here yields 4.4 times less false positive and 2.1 times less false negative detection compared to the traditional threshold method.

The PAL system is effective in the continuous observation of the presence of cloud hydrometeors. Together with a robust signal analysis technique, it proves to be an effective multipurpose remote sensing tool in observing the atmosphere. It is particularly superior in detecting optically thin high cirrus clouds over the observation site. Moreover, the interaction between aerosols and clouds is important in radiation budget studies, and the large PAL dataset can be used in cloud/aerosol transition.

Acknowledgment

This study was partly supported by the Natural Sciences Research Institute of the University of the Philippines.

References

[1] N. Lagrosas, Y. Yoshii, H. Kuze, N. Takeuchi, S. Naito, A. Sone, H. Kan. "Observation of boundary layer aerosols using a continuously operated, portable lidar system". Atmospheric Environment **2004**, 38, 3885-3892.

[2] N. Lagrosas, H. Kuze, N. Takeuchi, S. Fukagawa, G. Bagtasa, Y. Yoshii, S. Naito, M. Yabuki. "Correlation study between suspended particulate matter and portable automated lidar data". Journal of Aerosol Science **2005**, 36(4), 439-454.

[3] G. Bagtasa, N. Takeuchi, S. Fukagawa, H. Kuze, S. Naito. "Correction in aerosol mass concentration measurements with humidity difference between ambient and instrumental conditions". Atmospheric Environment **2007**, 41, 1616-1626.

[4] R.M. Measures. *Laser Remote Sensing: Fundamentals and Applications.* John Wiley & Sons: New York, USA, 1984.

[5] S. Lerkvarnyu, K. Deijhan, F. Cheevasuvit, "Moving average method for time series Lidar data". http://www.gisdevelopment.net/aars/acrs/1998/ps3/ps3016.asp, 1998.

[6] J.B. Buckheit, D.L. Donoho. Wavelab and Reproducible Research. http://www-stat.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf, 1995.

[7] C.S. Burrus, R.A. Gopinath, H. Guo. *Introduction to Wavelets and Wavelet Transforms, A PRIMER*. Prentice Hall: Upper Saddle River, NJ, USA, 1998.

[8] M. Lang, H. Guo, J.E. Odegard, C.S. Burrus, R.O. Wells Jr. "Noise reduction using an undecimated discrete wavelet transform", IEEE Signal Processing Letters **1996**, 3, 10-12.

[9] D.L. Donoho. "De-noising by soft-thresholding", IEEE Trans. Inform. Theory **1995**, 41, 613-627.

[10] H.T. Fang, D.S. Huang, "Noise reduction in lidar signal based on discrete wavelet transform". Optics Communication **2004**, 233, 67-76.

[11] C.M. Platt, S.A. Young, A.I. Carswell, S.R. Pal, M.P. McCormick, D.M. Winker, M. DelGuasta, L. Stefanutti, W.L. Eberhard, M. Hardesty, P.H. Flamant, R. Valentin, B. Forgan, G.G. Gimmestad, H. Jager, S.S. Khmelevtsov, I. Kolev, B. Kaprieolev, Da-ren Lu; K. Sassen, V.S. Shamanaev, O. Uchino, Y. Mizuno, U. Wandinger, C. Weitkamp, A. Ansmann, C. Wooldrigde. "The Experimental Cloud lidar Pilot Study (ECLIPS) for cloud radiation research". Bull. Amer. Meteor. Soc. **1994**, 75, 1635-1654.

[12] S.R. Pal, W. Steinbrecht, A. Carswell. "Automated method for lidar determination of cloud base height and vertical extent". Applied Optics **1992**, 31, 1488-1494.

[13] E.E. Clothiaux, G.G. Mace, T.P. Ackerman, T.J. Kane, J.D. Spinhirne, V.S. Scott. "An Automated Algorithm for Detection of Hydrometeor Returns in Micropulse Lidar Data". J. Atmos. Oceanic Technol. **1998**, 15, 1035–1042.

[14] K. Sassen, B.S. Cho. "Subvisual-thin cirrus lidar data set for satellite verification and climatological research". J. Appl. Eteor. **1992**, 31, 1275-1285.

[15] J.A. Walker, *Primer on Wavelets and Their Scientific Applications,* 2nd ed.; Chapman & Hall/CRC, Boca Raton, Florida, USA, 2008.

[16] F.X. Kneizys, E.P. Shettle, L.W. Abreu, J.H. Chetwynd Jr., G.P. Anderson, W.O. Gallery, J.E.A. Selby, S.A. Clough, R.W. Fenn. Users Guide to Lowtran7 1988. Air Force Geophysics Laboratory, Bedford, MA.